

“Tastes, ties, and time” through the lenses of Salganik’s Four Guiding Principles.

Big data has the potential to change the way social sciences are conducted. However, it also has the potential for ethical problems because the rudimentary understating on how to apply current ethical notions to research in the digital space. (Boyd and Crawford 2012) Salganik has proposed four principles to help to evaluate ethical issues of research in the digital age. Applying these principles to “Tastes, ties, and time” permitted me to get a more nuanced understanding about the ethical shortcomings of the study; as well as, helping me to decide if I would use the data for my research and shined a light on possible improvement to the ethics when using the data set.

Did the authors of “Tastes, Ties, and Time” follow the principle of Respect for Persons? Even when participants were not exposed to risk, manipulation, or deception; and researchers received the approval of Facebook and the Institutional Review Board (IRB) to access students’ profiles, the authors still failed to adhere to the principle of Respect for Persons, violating peoples’ autonomy and privacy when they did not ask for informed consent. (Zimmer 2010:320)

First, most social networks users do not read the Terms of Services. Furthermore, it is unclear that Facebook’s Terms of Service at the time allowed the company to use users’ data for the purpose of social research. Thus, it is not possible to convincingly argue that users consented to the study when they agreed to Facebook’s Terms of service. (Zimmer 2010:320)

Moreover, the researchers’ perception that the data was public and the approval of the IRB highlights a lack of information regarding ethical concerns and ideas emerging at the time such Contextual Privacy; notions that according to Salganik, could be derived from the Belmont Report. (Association of Internet Researchers n.d., Zimmer 2010:318) According to these ideas, privacy is a context-sensitive notion that depends on rules related to the amount and kind of information being accessed and the way information this is shared. (Zimmer 2010:322-323, Zimmer 2011, Nissebaum n.d., Boyd and Crawford 2012) Thus, even if the data was public, and users were bound by Facebook’s Terms of Service, according to Contextual Privacy, the authors might have violated users’ privacy when, without consent, collected users’ information yearly (amount of information), and created new kinds of information, non-shared in the Facebook profiles, merging the profiles data set with the administrative data set. (Lewis et al. 2008)

Additionally, because, the participants intended to restrict the visibility of their Facebook profiles to the Harvard College network (this is why Research Assistants with access to the college network were used to gain access to profiles) researchers might have violated the rules of contextual flow of information. When researchers shared the merged data set with investigators outside the Harvard College network, they broke the contextual restriction that students placed on their profiles. (Zimmer 2011)

Finally, while it is feasible that asking for Informed Consent could have introduced biases to the study, some kind of debriefing consent could have been used, especially considering the technology used and the small size of the data set, 1700 participants. Moreover, it appears, researchers did not offer additional protections to individual with limited autonomy.

On the other hand, it is important to take into consideration Salganik's warning of not allowing one principle to take precedence over others, but see them as components of a four-fold framework.

The research question and limitations of the available data at the time seem to justify the research. For example, it is clear that a more nuanced understanding of the balancing mechanism involved in the production of high degrees of racial homogeneity in social networks could deepen Social Science's understanding of the phenomenon while contributing to society's general understanding of factors related to segregation in everyday lives and into different social domains. (Wimmer and Lewis 2010:584-585)

In this sense, the characteristic of the data, even when the bias of selection made the data not generalizable to other populations, allowed a rigorous assessment of the process of racial homogeneity that has been of interest to social scientists for a long time. (Lewis, Gonzalez, and Kaufman 2012:62-63, Wimmer and Lewis 2010:584-585) Additionally, because the study was observational, the researchers did not place participant under direct risk. Similarly, because the necessary measures were taken to minimize informational risk (for example the process of anonymization) the study complied with notions of beneficence. (Zimmer 2010, Lewis et al. 2012:62)

On the other hand, the authors could have taken further measures to establish an adequate balance of risks and benefits. It is unclear if a thorough risk/benefit analysis was conducted and the researchers did an inadequate assessment of the potential risks known at the time.

While the potential risk to the Harvard College network, as a system it's hard to assess, the risks to the scientific community (such as loss of trust) and future users of the data was not adequately assessed. (Hatmaker 2014, Zimmer 2011) Similarly, the research created a precedent to studies that went from the observation to experimentation with Facebook users: for example, Facebook's Emotional Contagion study. (Hatmaker 2014)

Furthermore, the risk posed by the emergence of new information when data sets were combined and the limitations of anonymization were not completely evaluated. Researchers should have taken more drastic strategies regarding the anonymization process, especially when combining data sets that aggregate demographics and cultural traces. Especially, after cases such as the AOL one, where users were re-identified after the release of individuals' search history, and the release of studies ("Information Revelation and Privacy in Online Social Networks (The Facebook case)") assessing the weakness of anonymization. (Zimmer 2010:319, Boyd and Crawford 2012, Gross and Acquisti 2005)

The project adheres more closely to the principle of justice. Because the diversity of the Harvard College 2009 freshman class, individuals from underrepresented sectors were included in the research and potentially could access its benefits. Furthermore, since the study was observational, there was not abusive or excessive treatment by the researchers. (Wimmer and Lewis 2010:605) On the other hand, while the general community benefits from the research's outcomes, the burden of the study was placed only on the students' community of Harvard

College. Furthermore, the participant did not receive any kind of compensation, in monetary or educational form.

Finally, another guiding principle to which the project was more closely adhered is the Respect for Law and Public Interest. While the research project required research assistants, with access to the Harvard College Network, in order to circumvent the restrictions on the access to user's profiles, the authors complied with current laws and regulations such as Terms of Service, as indicated by the Facebook and IRB approval. (Lewis et al. 2008) Furthermore, under the requirement of the National Science Foundation, the researchers opened their data and methods to the public, avoiding secrecy, and adhering to the transparency-based Accountability.

Finally, regarding the question, if I would use the data, after careful consideration of the research's ethics, I conclude that I would use it for several reasons. First, while the project violated Salganik's Respect for Persons and poorly implemented the Beneficence principles, the data collection was conducted with not nefarious intentions and the ethics shortcomings, while not justifiable, can be understood because of the "murkiness" of the ethics in data science, in particular at the time. Finally, the characteristics of the data make it an ideal data set for researching how cultural diffuses through social networks, a research interest of mine.

Having said that, I will like to propose several improvements I will perform. I will try to develop stricter strategies for anonymization, especially in relation to the emergence of identifiable information after combining data sets; specifically, those related with demographic and cultural traces that could be used to identify participants. Finally, following Nissenbaum's notion of Contextual Privacy, I will try to establish clearer and more robust rules regarding informational flow. (Nissenbaum n.d., Boyd and Crawford 2012:673)

Citations

Association of Internet Researchers. n.d. Ethical Decision-Making and Internet Research

Boyd, Danah and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* 15(5):662–79.

Gross, Ralph and Alessandro Acquisti. 2005. "Information Revelation and Privacy in Online Social Networks."

Hatmaker, Taylor. 2014. "In 2006, Harvard Also Conducted a Facebook Study That Went Too Far ." *The Daily Dot*. Retrieved October 8, 2016 (<http://www.dailydot.com/debug/facebook-t3-study-tastes-ties-time/>).

Lewis, Kevin, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis. 2008. "Tastes, Ties, and Time: a New Social Network Dataset Using Facebook.com." *Social Networks* 30(4):330–42.

Lewis, Kevin, Marco Gonzalez, and Jason Kaufman. 2012. "Social Selection and Peer Influence in an Online Socialnetwork." Pp. 68–72 in, vol. 109.

Nissebaum, Helen. n.d. "Privacy as Contextual Integrity." Washington Law Review.

Wimmer, Andreas and Kevin Lewis. 2010. "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook." American Journal of Sociology 116(2):583–642.

Zimmer, Michael. 2010. "'But the Data Is Already Public': on the Ethics of Research in Facebook." Ethics and Information Technology 12(4):313–25.

Zimmer, Michael. 2011. "My Research in the Chronicle of Higher Education: 'Harvard's Privacy Meltdown'; Some Annotations ." Retrieved October 8, 2016 (<http://www.michaelzimmer.org/2011/07/11/harvards-privacy-meltdown/>).